

# Support Verb Constructions across the Ocean Sea

Jorge Baptista<sup>1,2</sup>, Nuno Mamede<sup>2,3</sup> & Sónia Reis<sup>1,2</sup>

<sup>1</sup>University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

<sup>2</sup>INESC-ID Lisboa - Human Language Technology Lab, Alves Redol 9, 1000-029 Lisboa

<sup>3</sup>Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais 1, 1049-001 Lisboa

[jbaptis@ualg.pt](mailto:jbaptis@ualg.pt), [nuno.mamede@tecnico.ulisboa.pt](mailto:nuno.mamede@tecnico.ulisboa.pt), [smreis@ualg.pt](mailto:smreis@ualg.pt)

# abstract

This paper analyses the *support (or light) verb constructions* (SVC) in a publicly available, manually annotated corpus of multiword expressions (MWE) in Brazilian Portuguese. The paper highlights several issues in the linguistic definitions therein adopted for these types of MWE, and reports the results from applying STRING, a rule-based parsing system, originally developed for European Portuguese, to this corpus from Brazilian Portuguese.

The goal is two-fold: to improve the linguistic definition of SVC in the annotation task, as well as to gauge the major difficulties found when transposing linguistic resources between these two varieties of the same language.

**Keywords:** support-verb constructions, light verbs, predicate noun, Portuguese, Brazilian/European Portuguese

# summary

**introduction:** concept of Support Verb Constructions (SVC), definitory properties, goals

**related work:**

**SVC in PARSEME:** LVC.full/cause, corpus compilation/annotation, lexical variety; linking operator-verb (Vopl), causative operator-verb (Vopc), standard/converse SVC

**processing SVC** in PARSEME corpus through STRING

**results and discussion:** FN: on-going/missing, missing Vsup, parsing issues:

TN: PT-BR exclusive SVC; FP: parsing issues

**conclusion and future work**

# introduction

## Support verb constructions (SVC)

- ❑ fundamental component of many languages (universal?)
- ❑ large variety of semantic predicates (like predicative verbs and adjectives)
- ❑ elementary (kernel) sentence (Z.S. Harris 1955, 1964, 1982, 1991)
- ❑ support verb + predicate noun + argument(s) (M. Gross 1981)
- ❑  $V_{sup}$  serves basically to “conjugate” the predicate noun (M.Gross 1989, p.38)
- ❑  $V_{sup}$  conveys grammatical (tense, aspect) and stylistic values (M.Gross 1998)

(terminology note: support verb = *light* verb; Fotopoulou *et al.* 2021)

## introduction (cont.)

- (1) O *Pedro* *tem* *fome*  
lit: 'Pedro has hunger' 'Pedro is hungry' (Santos 2015)
- (2) O *Pedro* *deu* *um* *abraço* *ao* *João*  
'Pedro gave a hug to João' (Baptista 1997b; Calcia 2022)
- (3) O *Pedro* *fez/está\_em* *greve*  
'Pedro is on strike' (Chacoto 2005; Dias de Barros 2014; Ranchhod 1990)
- (4) *É de\_o interesse do Pedro que o João faça isso*  
'It is in Pedro's interest that João do this' (Baptista 2005).

# introduction (cont.)

## Support verb constructions (SVC)

- ❑ SVC properties defined by V<sup>sup</sup>+N<sub>pred</sub> combination
- ❑ regularities can be found but can not be generalized over the lexicon of neither V<sup>sup</sup> nor N<sub>pred</sub>
- ❑ definitory properties, distinguishing SVC from ordinary verb constructions e.g. argument selection:

(5) *O Pedro<sub>i</sub> deu um abraço<sub>i</sub> (\*meu<sub>j</sub>/\*do Rui<sub>k</sub>) ao João*  
'Pedro gave a/my/Rui's hug to João'

# introduction (cont.)

This paper's goals:

- ❑ assess (on-going) SVC lexicon-grammar description of (European) Portuguese (PT-PT) (see references)
- ❑ assess its implementation in STRING, a hybrid, rule-based and statistical natural language processing pipeline for PT (Mamede *et al.* 2012, Baptista & Mamede 2020)
- ❑ use a publicly available, independently annotated, corpus (PARSEME) built for Brazilian Portuguese (PT-BR), with SVC manually annotated
- ❑ discuss several issues in the linguistic definition of SVC adopted in this corpus
- ❑ gauge the major difficulties when applying a lexicon-grammar built for PT-PT to a corpus of PT-BR, highlighting the similarity of the two major varieties of Portuguese

## related work

- quite old, the idea of *nouns as predicates* (Port-Royal, Arnaud & Lancelot 1660)
- modern thread: Zellig S. Harris (1955, and subsequent work: 1964, 1976, 1982, 1991)
- the terms *support verb* and *predicative noun* coined by M. Gross (1981)
- concepts later revised and extended (M. Gross 1998)
- extensive/systematic descriptions of SVC within the Lexicon-Grammar framework for major European and other non-IE languages (M. Gross 1996)
- European Portuguese (early 80's and 90's); Brazilian Portuguese (2013-2022)
- recent and brief, non-exhaustive overview on the topic (Fotopoulou *et al.* 2021)



# related work (cont.)



## ☐ CVS in Portuguese (overview)

### European Portuguese

- ☐ late **80s** (Vaza 1988: *dar* 'give'; Ranchhod 1990 [1988]: *estar Prep*) 'be Prep', late **90s** (Baptista 1997b: *dar/levar* 'give/take'), and early **2000s** (Baptista 2005a [2000]: *ser de* 'be of'; Chacoto 2005: *fazer* 'make/do')
- ☐ specific constructions and/or transformations: *conversion* (Baptista 1997a; Baptista 1997b, Baptista 2005a); *fusion* and instrument nouns (Baptista 2004); *symmetry* (Baptista 2005b), *communication* predicates (Reis et al. 2021), non-agentive SVC (Baptista et al. 2021)

### Brazilian Portuguese

- ☐ from 2010s (Dias de Barros 2014: *fazer*; Rassi 2015: *dar* 'give'; Santos 2015: *ter* 'have')
- ☐ specific constructions and/or transformations: *conversion* (Calcia 2016, Calcia and Vale 2019; Calcia 2022), aspectual variants of *vsup* (Picoli et al. 2021), *non-agentive SVC* (Dias de Barros et al. 2013)

## related work (cont.)



- ❑ Few works systematic comparison of the lexicon and grammar of the PT-PT and PT-BR variants (Rassi *et al.* 2016: converse SVC)
- ❑ annotated corpus of SVC with support-verb *dar* 'give' (Rassi *et al.* 2015b)
- ❑ processing SVC in Portuguese (Baptista *et al.* 2015; Rassi *et al.* 2014; Rassi *et al.* 2015a)
- ❑ recent developments in (Mota *et al.* 2018; Baptista & Mamede 2020; Barreiro *et al.* 2022)
- ❑ extensive literature on SVC across multiple languages, on their place within the description of multiword expressions (Sag *et al.* 2002); relation of SVC with fixed, verbal idioms; challenges SVC pose to Natural Language Processing (NLP) (Constant *et al.* 2017, for an overview)
- ❑ significant boost multilingual community COST project PARSEME (Savary *et al.* 2015)

# related work (cont.)

## PARSEME project

- ❑ goal “characterizing MWEs in lexicons, grammars and corpora and enabling systems to process them” (Ramisch *et al.* 2020, p.107)
- ❑ shared task on automatic identification of MWE (Ramisch *et al.* 2020)
- ❑ (Brazilian) Portuguese corpus, manually annotated for verbal MWE
- ❑ “unified guidelines for all the participating languages, in order to avoid heterogeneous, hence incomparable, datasets”: decision-trees
- ❑ main categories: SVC (called LVC=light verb) and verbal idioms

# SVC in PARSEME

Some data (from documentation)\* :

- ❑ 80% training + 10% development + 10% *testing* partitions
- ❑ total: 27,9 K sentences, 638 K tokens
- ❑ 3,145 **LVC.full** (= SVC) and 94 **LVC.cause** (=causative constructions)
- ❑ *testing* partition: 2,770 sentences, +62.6 K tokens, 337 LVC.full, 7 LVC.cause
- ❑ “sentences from the informal Brazilian newspaper Diário Gaúcho and from the training set of the [Universal Dependencies] treebank” (UD\_Portuguese-GSD v2.1); no further information found on the sampling/selection method for the sentences included in the corpus
- ❑ 1,000 sentences doubly-annotated; agreement reported seems to indicate adequate guidelines, yielding reasonable consistency, given the complexity of the task

\* Ramisch et al. (2020); [https://gitlab.com/parseme/parseme\\_corpus\\_pt](https://gitlab.com/parseme/parseme_corpus_pt)

# SVC in PARSEME (cont.)

## ❑ Lexical variety

(for the purpose of this paper)

- ❑ football (soccer) domain vocabulary:
  - ❑ 134 instances of **gol** 'goal' (*fazer* 'do' and *marcar* 'score', ?*dar* 'give');
  - ❑ 30 *falta* (*fazer*, *marcar*, *sofrer* 'suffer', ?*cometer* 'commit');
  - ❑ 13 *passe* (*fazer* 'make, do', *receber* 'receive');
  - ❑ 11 *pênalti* 'penalty' (*marcar*, *sofrer*); etc.

**6.8% SVC** instances concern vocabulary of this domain

## SVC in PARSEME (cont.)

- predicate nouns designating **measurable quantities**:

8 *área* 'area', 2 *população* 'population'; some taken from the Astronomy domain: 2 *ascensão reta* 'right ascension', 2 *declinação* 'declination', 6 ***excentricidade*** 'eccentricity', ***inclinação*** 'inclination':

*Possui uma **excentricidade** de 0.03574140 e uma **inclinação** de 11.03095°.*

'It has an eccentricity of 0.03574140 and a tilt of 11.03095°.'

## SVC in PARSEME (cont.)

- ❑ predicate nouns designating **measurable quantities**:  
*Possui uma **excentricidade** de 0.03574140 e uma **inclinação** de 11.03095°.*  
'It has an eccentricity of 0.03574140 and a tilt of 11.03095°.'
- ❑ apparently, no occurrence of other SVC of some of these nouns (Baptista 2005):  
*O Pedro é de uma certa **excentricidade***  
lit: 'Pedro is of a certain eccentricity' ('Pedro is eccentric')  
*O Pedro tem uma certa **inclinação** para a música*  
lit: 'Pedro has a certain inclination to music ('Pedro ')
- ❑ nor any case of **Restructuring** (Baptista & Ranchhod, 1998):  
*O mundo tem 510 bilhões de km<sup>2</sup> de **área** total* ([www.ufjf.br](http://www.ufjf.br))  
'The world has 510 billion km<sup>2</sup> of total area'

## SVC in PARSEME (cont.)

- Linking operator verbs (M. Gross 1981)

*A Cátedra Milton Santos **tem como objetivo** a difusão de informações* (id=pt  
br-ud-train-s7942)

‘The Milton Santos Chair has as its aim/purpose to disseminate information’

*o especial da TV Globo **terá como tema** a vida de Dolores Duran* (id=diario gauchó 16311)

‘the TV Globo special will have as theme the life of Dolores Duran’



## SVC in PARSEME (cont.)

- there is an underlying SVC, with standard syntactic configuration, and the predicative natures of *como* hints at the sentence with copula verb:

*A (constituição da) Cátedra Milton Santos **tem** um certo **objetivo***

*# esse objetivo **é** a difusão de informações*

'The (constitution of the) Milton Santos Chair has as an aim/purpose

# This aim/purpose is to disseminate information'

*o especial da TV Globo **terá** um certo **tema** # o tema **é** a vida de Dolores Duran*

'the TV Globo special will have a theme # This theme is the life of Dolores Duran'

## SVC in PARSEME (cont.)

- Linking operator verbs (continued), on adjectival/participial constructions:

*Já Federer . . . **teve** uma **campanha** mais **perturbada** ...*

'Federer [=person] ... had a more troubled/disrupted campaign'

cp. *a **campanha** (de Federer) foi **perturbada*** 'the campaign was troubled/disrupted'

## SVC in PARSEME (cont.)

- Linking operator verbs (continued), on prepositional constructions:

*Pelo segundo ano consecutivo, o Cruzeiro **teve** uma **campanha abaixo de**\_as expectativas.*

(lit.: had a campaign *below* expectations')

'For the second year in a row, Cruzeiro [football club] has fallen short of expectations'

cp. *a **campanha** (do Cruzeiro) esteve **abaixo de**\_as expectativas*

'the campaign was below expectations'

## SVC in PARSEME (cont.)

- ❑ A clear SVC construction of the noun **campanha** ‘campaign’ is patent in the corpus:

*Os jovens...estão **fazendo** a **campanha** com a cara e a linguagem deles*

‘Young people are making the campaign with their face and their language’

- ❑ some cases where the notion of support seems too much stretched:

*... o partido foi vítima de uma intensa **campanha promovida** pela oposição de direita e seus aliados ...* ‘the party was the victim of an intense campaign promoted by the right-wing opposition and its allies’

cp. *O Pedro promoveu a **campanha** (do João = que o João **fazia**)*

‘Pedro promoted the campaign (of João/João’s = that João was doing)’

- ❑ *ser vítima de Npred* and SVC with copula (see below)

## SVC in PARSEME (cont.)

- ❑ Causative operator verbs (Vopc; M. Gross 1981),  
construct integrated in PARSEME Guidelines (LVC.cause)

*A ausência do sexo também **traz** uma forte **angústia*** (OK: marked as LVC.cause)

'The absence of sex also brings a strong anguish'

*Nós ... estamos ansiosos para montar um time competitivo, que seja divertido e **traga orgulho** para os fãs* (Vopc was missed)

'We are looking forward to building a competitive team that is fun and brings pride to the fans'

*Caro V., a tua postura é sempre admirável, o que **faz com que** tua lealdade a esta coluna só me **dê orgulho*** (one Vopc missed and another marked as LVC.full)

'Dear V., your posture is always admirable, which makes me proud of your loyalty to this column'

# SVC in PARSEME (cont.)

## operator-verbs (conclusion)

- ❑ distinction between **Vsup** and **Vopc** construction is not consensual
  - ❑ (e.g. PT-PT: de Athayde 2000 vs. Chacoto 2005)
  - ❑ Gross (1981) original proposal vs. Gross (1998) revised as type of SVC
  - ❑ this paper concurs with PARSEME classification criteria:
    - semantic neutral value of Vsup vs. semantic added value of **cause** of Vopc
- ❑ theoretical construct of linking operator-verb
  - ❑ also proposed by M. Gross (1981), but largely ignored in the literature
  - ❑ treated as a ordinary Vsup in PARSEME (or just ignored)
  - ❑ no new meaning element is introduced, came predicate noun (and partially its structure)
  - ❑ proposed as a **saliency-inducing device** (Ranchhod 1990, 183 ff.)
  - ❑ a new notation is required, distinct from Vsup and Vopc



## SVC in PARSEME (cont.)

### ❑ **Standard/Converse SVC; [Conversion]** (M. Gross 1981; G. Gross 1989)

(ST)      *O Pedro deu um abraço ao João*      'Peter gave John a hug'  
= (CV)    *O João recebeu/levou um abraço do Pedro*    'John got a hug from Peter'

- ❑ very productive transformation, occur in several (many) languages
- ❑ extensive descriptions in PT-PT and PT-BR  
(Baptista 1997a, 1997b, 2005b; Rassi et al. 2016; Calcia 2016; Calcia & Vale 2019; Calcia 2022).
- ❑ occurred often in the corpus (34 times)
- ❑ may have an impact in the semantic role labelling
- ❑ since PARSEME ignored this distinction,  
this Standard/Converse opposition was also ignored



# processing SVC in PARSEME corpus with *String*

- ❑ STRING pipeline (Mamede *et al.* 2012; Baptista & Mamede 2020)
  - ❑ basic text processing tasks, including text segmentation, tokenization, dictionary-based part-of-speech (PoS) tagging, rule-based and statistical PoS disambiguation, and parsing.
  - ❑ XIP Parser (Ait-Mokhtar *et al.* 2002) with the STRING Portuguese lexicons and grammar
    - Dependency relations: SUBJ[ect], CDIR (direct object), etc., between chunks heads
  - ❑ SVCs are formally identical to ordinary verbal constructions > STRING overall strategy
    - ❑ *firstly*, capturing (general) syntactic dependencies between predicate noun and the support verb; and
    - ❑ *then* extracting a specific dependency SUPPORT linking the predicate noun and the support verb.
  - ❑ The system can be configured to output only the desired dependencies.

# processing SVC in PARSEME corpus with STRING (cont.)

*A Ana marcou dois gols* ‘Ana has scored (lit.: marked) two goals’

>TOP{NP{A Ana} VF{marcou} NP{dois gols}}

SUBJ\_PRE(marcou,Ana)

CDIR\_POST(marcou,gols)

SUPPORT\_VSUP-STANDARD(gols,marcou)

*String*

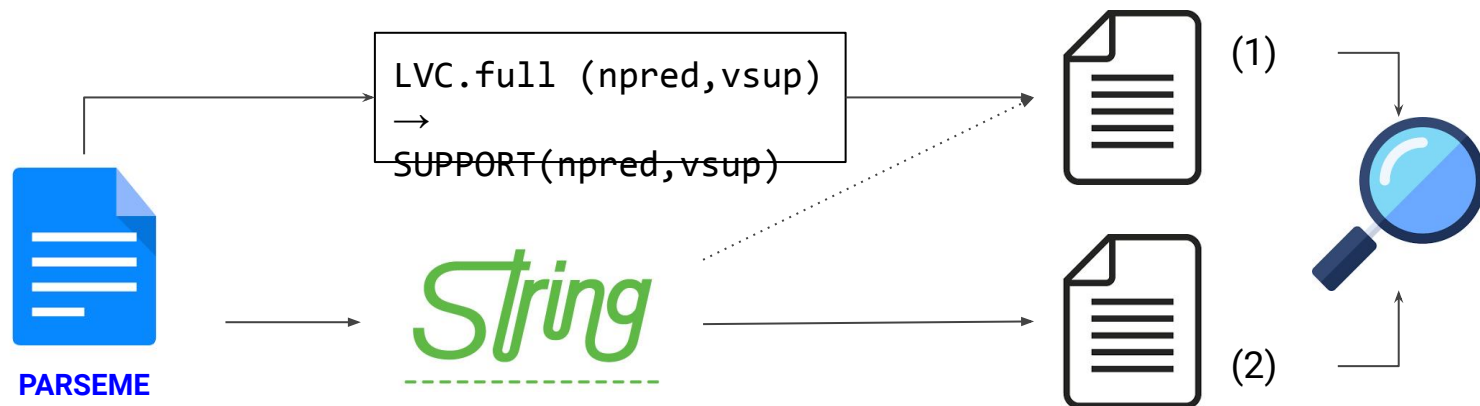
# processing SVC in PARSEME corpus with STRING (cont.)

```
if ( ( VDOMAIN(#1,#2[lemma:"fazer"]) || VDOMAIN(#1,#2[lemma:"marcar"]) )
    & ( MOD[post,relat](#3[lemma:"golo"],#2) ||
        CDIR(#2[transf-passiva:~],#3[lemma:"golo"]) ||
        SUBJ(#2[transf-passiva],#3[lemma:"golo"]) ||
        ( ANTECEDENT[relat](#3[lemma:"golo"],#4[pronrel])
          & SUBJ(#2[transf-passiva],#4)))
    & ~ SUPPORT[vsup-standard](#3,#2) )
SUPPORT[vsup-standard=+](#3,#2)
```

- ❑ **15,615 rules** automatically generated from the database (lexicon-grammar; Baptista & Mamede 2020, *updated*)
- ❑ **5,800 SVCs (+3,320 predicate nouns, ongoing)**;
- ❑ gauge lexical coverage of the resource and the system's performance on a previously unseen corpus, from a different variety

The logo for the STRING project, featuring the word "String" in a green, cursive font with a dashed underline.

# processing SVC in PARSEME corpus with STRING (cont.)



- ❑ (in-house script) retrieve {npred, vsup} pairs linked by LVC.full dependency from corpus
- ❑ corpus processed by STRING, (only) with SUPPORT dependency output: (2)
- ❑ convert LVC.full into SUPPORT and merge with processed corpus output (1)
- ❑ semi-automatic comparison

# processing SVC in PARSEME corpus with STRING (cont.)

1>TOP{PP{PP{Em 2} PP{de outubro} PP{de 2009}},  
PP{ em o 10º aniversário} PP{de a SmackDown},  
NP{The Rock} VF{**fez**} NP{uma **aparição**}  
AP{especial} PP{em um vídeo} AP{pré-gravado} .}

**PARSEME:** SUPPORT\_VSUP-STANDARD(aparição,fez)  
**STRING:** SUPPORT\_VSUP-STANDARD(aparição,fez)  
(true-positive)

2>TOP{NP{Se} NP{nenhuma **providência**}  
VCOP{for} VCPART {**tomada**}, NP{a população}  
VTEMP{vai} VASP{voltar a} VINFIN{usar}  
NP{lamparinas} ADVP{ADV{a a noite}} e NP  
{geladeira} PP{a querosene} .}

**PARSEME:** SUPPORT\_VSUP-STANDARD(providência,tomada)  
**STRING:** -- (false-negative)



# results and discussion



- ❑ 2,770 sentences (as segmented by STRING)
- ❑ 311 instances of LVC.full in PARSEME
- ❑ high precision / low recall

## Error analysis

- ❑ false-negatives:
  - ❑ ongoing predicate nouns' description
  - ❑ missing nouns
  - ❑ missing support verb in the lexicon-grammar
  - ❑ pos-tagging and parsing issues
- ❑ true-negatives
  - ❑ predicate nouns *exclusive* of PT-BR
  - ❑ lexical variants *exclusive* of PT-BR
  - ❑ spelling errors/variants
- ❑ false-positives
  - ❑ pos-tagging and parsing issues

<b>TP</b>	<b>FP</b>	<b>FN</b>
197	20	270
<b>P</b>	<b>R</b>	<b>F</b>
0.91	0.42	0.58

# results and discussion (cont.)

## □ ongoing predicate nouns' description

*baixa* 'sick leave', *convenção* 'convention', *convivência* 'coexistence', *dano* 'damage', *decisão* 'decision', *dever* 'duty', *disponibilidade* 'availability', *extorsão* 'extortion', *facilidade* 'ease', *grandeza* 'greatness', *hábito* 'habit', *maneira* 'manner', *medida* 'measure', *obra* 'work', *perda* 'loss', *rachadura* 'crack', *reclamação* 'claim', and *validade* 'validity'

## □ missing nouns

*aniversário* 'anniversary', *antecedentes* 'background', *área* 'area', *autonomia* 'autonomy', *característica* 'features', *chefe* 'chief', *crime* 'crime', *endereço* 'address', *equivoco* 'equivocation', *êxito* 'success', *favoritismo* 'favouritism', *grafitagem* 'graffiti', *homicídio* 'homicide', *índice* 'index', *lar* 'home', *lembrança* 'memory', *moleza* 'softness/sloth', *padrão* 'pattern', *passado* 'past', *população* 'population', *potencial* 'potential', *prazer* 'pleasure', *presença* 'presence', *problema* 'problem', *procedimento* 'procedure', *propriedade* 'property', *repertório* 'repertoire', *significado* 'significance', *subvenção* 'grant', *tempo* 'time', *tratado* 'treaty', *treinamento* 'training', *vínculo* 'bond'; *matchpoints* (idem), PT-BR: *turnê* / PT-PT: *turné* (from French: *tournee*'), *video-chamada* / *videochamada* (ortographic variant)

## results and discussion (cont.)

- ❑ **missing support verb** in the predicate noun entry of the lexicon-grammar  
**realizar** 'do/make/perform': *ação* 'action', *apresentação* 'presentation', *audiência* 'audience', *concorrência* 'concurrency'; **cometer** 'commit': *assalto* 'burglary'; **assinar** 'sign': *acordo* 'agreement', *contrato* 'contract'; **ter** 'have': *cura* 'cure', *marcação* 'appointment', *relação* 'relationship'; **possuir** 'possess': *excentricidade* 'eccentricity' (Astron.), *experiência* 'experience', *inclinação* 'inclination' (Astron.), *poder* 'power'; **apresentar** 'present': *sinal* 'signal'

### converse support-verbs

**sofrer** 'suffer': *acidente* 'accident'; **levar** 'take': *advertência* 'warning', *medo* 'fear' (only in PT-BR), *tombo* 'fall'; **passar por** 'pass by'(=sofrer): *cirurgia* 'surgery'; **tomar** 'take': *cuidado* 'care/beware of', *gol* (PT-BR), *precaução* 'precaution', *providência* 'make arrangements for'; **chegar a** 'get to': *orgasmo* 'orgasm'



# results and discussion (cont.)

## ❑ parsing issues

- ❑ error in PoS-disambiguation: *ajuste* (noun/verb)

*A prática de **fazer ajuste** no superávit com os dividendos tem sido comum nos últimos anos*  
'The practice of **making adjustment** the surplus with dividends has been common in recent years'

- ❑ particular syntactic constructions not captured by the grammar

*Três integrantes de um bando que **fez um dos maiores ataques** a banco dos últimos anos [...]* 'Three members of a gang that **carried out one of the biggest bank attacks** in recent years'

complex *partitive determiner* construction (M. Gross 1977):

*fez **um** [ataque (que foi) ] **dos** maiores ataques*

only partially parsed (no CDIR): QUANTD\_PRE(ataques, um)

## results and discussion (cont.)

### ❑ false-positives

*Seria uma boa surpresa e uma **prova** de que amor não **tem** hora nem dia marcados.*  
'It would be a nice surprise and proof that love has no set time or day.'

**STRING: que** (PoS-tagging error: \*Pron-rel vs. Conj-sub)

INTROD\_RELAT(que,tem), ANTECEDENT\_RELAT(prova,que), MOD\_POST\_RELAT(prova,tem)  
→ SUPPORT\_VSUP-STANDARD(prova,tem)

SVC and the last stages of parsing, accumulation errors (PoS-tagging, parsing)

# conclusion and future work

This paper:

- ❑ analyzed the support-verb constructions (SVC) **manually** and **independently annotated** in a **publicly available** corpus of Brazilian Portuguese (PT-BR), originally built within the scope of the project PARSEME
- ❑ corpus was parsed with STRING NLP pipeline for Portuguese
- ❑ using PT-PT SVC lexicon-grammar (still on construction: 2/3)
- ❑ described SVC found in corpus
- ❑ compared SVC dependencies annotated in corpus with STRING system output
- ❑ high precision (.91) but low recall (.42), still much room for improvement
- ❑ error analysis: on-going (2/3), PT-BR nouns, PT-BR-specific or missing support-verb+noun combinations (e.g. *possuir* 'possess', *passar* 'pass'), PoS-tagging and parsing errors; still rule improvement
- ❑ (partition of) corpus with reference output available at:

*Thank you!*

# references

- Ait-Mokhtar, S., Chanod, J.-P., and Roux, C. (2002). Robustness beyond shallowness: Incremental deep parsing. *Natural Language Engineering*, 8, pages 121–144.
- Baptista, J. and Mamede, N. (2020). Syntactic Transformations in Rule-Based Parsing of Support Verb Constructions: Examples from European Portuguese. In Alberto Simões, et al., editors, 9th Symposium on Languages, Applications and Technologies (SLATE 2020), volume 83 of OpenAccess Series in Informatics (OASICs), pages 11:1–11:14, Dagstuhl, Germany. Schloss Dagstuhl– Leibniz-Zentrum für Informatik.
- Baptista, J. and Ranchhod, E. (1998). Propriétés des phrases élémentaires associées à l'expression de grandeurs mesurables. Exemples du portugais. *Cahiers de l'Institut de Linguistique de Louvain*, 24(3-4):49–61, September 24-27, 1998.
- Baptista, J., Rassi, A. P., Santos-Turati, C., de Barros, C. D., Vale, O. A., and Mamede, N. (2015). Integrated processing of support verb constructions in Portuguese, 19-20 March 2015 (PARSEME Workshop, Malta).
- Baptista, J. (1997a). Conversão, nomes parte-do-corpo e reestruturação dativa. In Ivo Castro, editor, *Actas do XII Encontro da Associação Portuguesa de Linguística*, volume 1, pages 51–59, Lisboa, 30 de Setembro a 2 de Outubro de 1996, Braga-Guimarães, Portugal. Associação Portuguesa de Linguística, APL/Colibri.
- Baptista, J. (1997b). 'Sermão', 'tareia' e 'fachada': uma classificação das expressões conversas 'dar-levar'. *Seminários de Linguística* 1, pages 5–37.
- Baptista, J. (1999). 'Fazer/Fazer com': Um verbo operador do português. *Seminários de Linguística*, 3:163–171.
- Baptista, J. (2004). Instrument nouns and fusion. predicative nouns designating violent actions. In Christian; Leclère, et al., editors, *Léxique, Syntaxe et Lexique-Grammaire (Syntax, Lexis and Lexicon-Grammar)*. Hommage a Maurice Gross. *Linguisticae Investigationes Supplementa*, pages 31–40. John Benjamins Publishing Co., Amsterdam/Philadelphia.
- Baptista, J. (2005a). Construções simétricas: argumentos e complementos. In Olga Figueiredo, et al., editors, *Estudos de homenagem a Mário Vilela*, pages 353–367. Faculdade de Letras da Universidade do Porto.
- Baptista, J. (2005b). *Sintaxe dos Predicados Nominais com 'ser de'*. Fundação para a Ciência e a Tecnologia & Fundação Calouste Gulbenkian, Lisboa.
- Barreiro, A., Mota, C., Baptista, J., Chacoto, L., and Carvalho, P. (2022). Linguistic resources for paraphrase generation in Portuguese: a lexicon grammar approach. *Language Resources and Evaluation*, pages 1–35.
- Borillo, A. (1971). Remarques sur les verbes symétriques. *Langue Française*, 11(1):17–31.
- Calcia, N. P. and Vale, O. (2019). Construções conversas do português do Brasil. *Descrição e classificação iniciais*. *Linguística*, 10(2):13–20, January.
- Calcia, N. (2016). *Descrição e classificação das construções conversas no português do Brasil*. Master's thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil.
- Calcia, N. P. (2022). 'Dar e receber um abraço': uma análise da conversão em português brasileiro. Ph.D. thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil (março).
- Chacoto, L. (2005). *O Verbo FAZER em Construções Nominais Predicativas*. Ph.D. thesis, Universidade do Algarve, Faro, Portugal.
- Constant, M., Eryigit, G., Monti, J., van der Plas, L., Ramisch, C., Rosner, M., and Todirascu, A. (2017). Multiword expression processing: A survey. *Computational Linguistics*, 43(4):837–892.
- Cook, P., Mitrovic, J., Escartin, C. P., Vaidya, A., Osen ova, P., Taslimipoor, S., and Ramisch, C. (2021). Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021). In *Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)*.
- de Athayde, M. F. M. Q.-P. (2000). *A estrutura semântica das construções com verbo-suporte preposicionadas do português e do alemão*. Ph.D. thesis, Faculdade de Letras da Universidade de Coimbra, Coimbra.
- Dias de Barros, C., Vale, O. A., and Baptista, J. (2013). Description and classification of nominal predicates with the support verb 'fazer' (make/do) in Brazilian Portuguese. In Jorge Baptista et al., editors, *Proceedings of the 32nd International Conference on Lexis and Grammar (CLG'2013)*, pages 165–170, Faro, Portugal, September. Universidade do Algarve – FCHS.
- Dias de Barros, C. (2014). *Descrição e classificação de predicados nominais com o verbo-suporte 'fazer': especificidades do Português do Brasil*. Ph.D. thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil.
- Fotopoulou, A., Laporte, E., and Nakamura, T. (2021). Where do aspectual variants of light verb constructions belong? In *Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)*, pages 2–12. Association for Computational Linguistics.
- Giry-Schneider, J. (1978). *Les nominalisations en Français: l'opérateur 'faire' dans le lexique*. Librairie Droz, Genova.

# references

- Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages*, 15(63):7–52.
- Gross, G. (1989). Les construction converses du français. Droz, Geneva.
- Gross, M. (1996). Lexicon-grammar. In Keith Brown et al., editors, *Concise Encyclopedia of Syntactic Theories*, pages 244–259. Pergamon, Cambridge.
- Gross, M. (1998). La fonction sémantique des verbes supports. *Travaux de Linguistique: Revue Internationale de Linguistique Française*, 37(1):25–46.
- Harris, Z. S. (1955). Co-occurrence and transformation in linguistic structure, pages 143–210. D. Reidel Publishing Company.
- Harris, Z. (1964). The elementary transformations. In Henry Hiz, editor, *Papers on Syntax*, pages 211–235. D. Reidel Publishing Company.
- Harris, Z. (1976). *Notes du Cours de Syntaxe*. Editions du Seuil, Paris.
- Harris, Z. S. (1982). *A Grammar of English on Mathematical Principles*. New York: John Wiley & Sons.
- Harris, Z. S. (1991). *A Theory of Language and Information. A Mathematical Approach*. Clarendon Press, Oxford.
- Mamede, N., Baptista, J., Diniz, C., and Cabarrão, V. (2012). STRING - A Hybrid Statistical and Rule-Based Natural Language Processing Chain for Portuguese. In Alberto Abad, editor, *Int. Conf. on Computational Processing of Portuguese (PROPOR 2012) - Demo Session*. <http://www.inesc.id.pt/ficheiros/publicacoes/8578.pdf>.
- Mota, C., Baptista, J., and Barreiro, A. (2018). The lexicon-grammar of predicate nouns with 'ser de' in port4nooj. In *International Conference on Automatic Processing of Natural-Language Electronic Texts with NooJ*, pages 124–137. Springer.
- Paiva Raposo, E. (2013). Predicações secundárias. In Paiva Raposo, et al., editor, *Gramática do Português*, volume 2, pages 1340–1356. Fundação Calouste Gulbenkian.
- Picoli, L., Vale, O. A., and Laporte, E. (2021). Aspecto verbal nas construções com verbo-suporte. *Revista do GEL*, 18(1):204–229.
- Ramisch, C., Savary, A., Guillaume, B., Waszczuk, J., Candito, M., Vaidya, A., Barbu Mititelu, V., Bhatia, A., Inurrieta, U., Giouli, V., Gungor, T., Jiang, M., Lichte, T., Liebeskind, C., Monti, J., Ramisch, R., Stymne, S., Walsh, A., and Xu, H. (2020). Edition 1.2 of the PARSEME shared task on semi-supervised identification of verbal multiword expressions. In *Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons*, pages 107–118, online, December. Association for Computational Linguistics.
- Ranchhod, E. (1990). *Sintaxe dos predicados nominais com 'estar'*. Instituto Nacional de Investigação Científica (INIC).
- Rassi, A., Santos-Turati, C., Baptista, J., Mamede, N., and Vale, O. (2014). The fuzzy boundaries of operator verb and support verb constructions with dar “give” and ter “have” in Brazilian Portuguese. In *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014)*, COLING 2014, Dublin, August 24, 2014, COLING 2014. Rassi, A., Mamede, N., Baptista, J., and I. O. V. (2015a). Integrating support verb constructions into a parser. In *Proceedings of the Symposium in Information and Human Language Technology (STIL'2015)*, pages 57–62.
- Rassi, A. P., Baptista, J., and Vale, O. A. (2015b). Um corpus anotado de construções com verbo-suporte em português. *Gragoata*, 39(1):207–230, Junho, 2015.
- Rassi, A., Calcia, N., Vale, O. A., and Baptista, J. (2016). Estudo contrastivo sobre construções com verbos em PB e PE. In O. Nadin, et al., editors, *Lexico e suas interfaces: descrição, reflexão e ensino*, pages 199–218. Cultura Acadêmica.
- Rassi, A. P. (2015). *Descrição, classificação e processamento automático das construções com o verbo dar em português brasileiro*. Ph.D. thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil.
- Reis, S., Mamede, N., and Baptista, J. (2021). Predicados de comunicação em português europeu: nominalizações e nomes autônomos. *Revista da Associação Portuguesa de Linguística*, (8):237–259, Out.
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002). Multiword Expressions: A Pain in the Neck for NLP. In Alexander Gelbukh, editor, *Proceedings of CILing*, pages 1–15, Mexico City, Mexico, February.
- Santos, C. (2015). *Construções com verbo-suporte 'ter' no Português do Brasil*. Ph.D. thesis, Universidade Federal de São Carlos, São Carlos-SP, Brasil.
- Savary, A., Sailer, M., Parmentier, Y., Rosner, M., Rosen, V., Przepiorkowski, A., Krstev, C., Vincze, V., Wojtowicz, B., Losnegaard, G. S., Parra Escartin, C., Waszczuk, J., Constant, M., Osenova, P., and Sangati, F. (2015). PARSEME-PARSING and Multiword Expressions within a European multilingual network. In *7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC 2015)*.
- Vaza, A. (1988). *Estruturas com nomes predicativos e verbo-suporte 'dar'*. Master's thesis, Faculdade de Letras da Universidade de Lisboa, Lisboa, Portugal.